



US Patent & Trademark Office

[Subscribe \(Full Service\)](#) [Register \(Limited Service, Free\)](#) [Login](#)

 Search: ☒ The ACM Digital Library ☒ The Guide

THE ACM DIGITAL LIBRARY


[Feedback](#) [Report a problem](#) [Satisfaction survey](#)
Terms used **outlier detection data mining**

Found 1,954 of 132,857

Sort results
by
[Save results to a Binder](#)Try an [Advanced Search](#)Display
results
[Search Tips](#)Try this search in [The ACM Guide](#)
☐ Open results in a new
window
Results 161 - 180 of 200 Result page: [previous](#) [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [next](#)

Best 200 shown

Relevance scale ☐ ☐ ☐ ☐ ☐**161** [Statistics and data mining techniques for lifetime value modeling](#)

D. R. Mani, James Drew, Andrew Betz, Piew Datta

August 1999 **Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining**Full text available: pdf(1.16 MB) Additional Information: [full citation](#), [references](#), [citations](#), [index terms](#)**162** [Data mining by business users: integrating data mining in business processes](#)

Marcel Holsheimer

August 1999 **Tutorial notes of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining**Full text available: pdf(1.72 MB) Additional Information: [full citation](#), [index terms](#)**163** [Towards automated synthesis of data mining programs](#)

Wray Buntine, Bernd Fischer, Thomas Pressburger

August 1999 **Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining**Full text available: pdf(637.67 KB) Additional Information: [full citation](#), [references](#), [index terms](#)**164** [Scalable algorithms for mining large databases](#)

Rajeev Rastogi, Kyuseok Shim

August 1999 **Tutorial notes of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining**Full text available: pdf(4.11 MB) Additional Information: [full citation](#), [references](#), [index terms](#)**165** [Invited talk: Untangling text data mining](#)

Marti A. Hearst

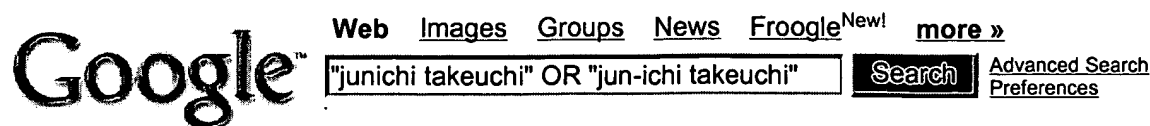
June 1999 **Proceedings of the 37th conference on Association for Computational Linguistics**

Full text available: pdf(781.57 KB)

Additional Information: [full citation](#), [abstract](#), [references](#)[Publisher Site](#)**166** [Verification and normalization of sentences](#)

Marco Costa, José Neves, Orlando Sousa, Simas Santos

June 1999 **Proceedings of the seventh international conference on Artificial intelligence and law**Full text available: pdf(237.35 KB) Additional Information: [full citation](#), [references](#), [index terms](#)

**Web**

Results 1 - 10 of about 561 for "junichi takeuchi" OR "jun-ichi takeuchi". (0.20 seconds)

DBLP: Jun-ichi Takeuchidblp.uni-trier.de **Jun-ichi Takeuchi**. ... 2003. 10, EE, Satoshi Morinaga, Kenji Yamanishi,**Jun-ichi Takeuchi**: Distributed cooperative mining for information consortia. ...www.informatik.uni-trier.de/~ley/db/indices/ a-tree/t/Takeuchi:Jun=ichi.html - 8k - Apr 28, 2004 - [Cached](#) -[Similar pages](#)**DBLP: Naoki Abe**... Computer Networks 33(1-6): 775-788 (2000). 33, **Jun-ichi Takeuchi**, Naoki Abe, Shun-ichi

Amari: The Lob-Pass Problem. J. Comput. Syst. Sci. 61(3): 523-557 (2000). ...

www.informatik.uni-trier.de/~ley/db/ indices/a-tree/a/Abe:Naoki.html - 18k - [Cached](#) - [Similar pages](#)[[More results from www.informatik.uni-trier.de](#)]**Jun-ichi Takeuchi****Jun-ichi Takeuchi**, Dr.Eng. Principal Researcher in Data Mining Research

Group Internet Systems Research Laboratories, NEC Corporation ...

www.labs.nec.co.jp/DTmining/members/ takeuchi/index_e.html - 2k - Apr 29, 2004 - [Cached](#) - [Similar pages](#)[[More results from www.labs.nec.co.jp](#)]**Jun'ichi Takeuchi's Page**

竹内純一. 博士(工学) NEC インターネットシステム研

究所 データマイニングリサーチグループ 主任研究

員 ...

www.labs.nec.co.jp/DTmining/members/ takeuchi/index_j.html - 2k - [Cached](#) - [Similar pages](#)**On-line unsupervised outlier detection using finite mixtures with ...**

Google, Inc. Subscribe (Full Service), Register (Limited Service,

Free), Login. Search: The ACM Digital Library The Guide. ...

portal.acm.org/ citation.cfm?id=347160&coll=portal&dl=ACM&CFID=11111111&CFTOKEN=2222222 -

[Similar pages](#)**Discovering outlier filtering rules from unlabeled data**... 216-8555, Japan. **Jun-ichi Takeuchi**, NEC Corporation, 4-1-1, Miyazaki, Miyamae,

Kawasaki, Kanagawa 216-8555, Japan. Sponsors, SIGMOD: ACM ...

portal.acm.org/ citation.cfm?

id=502570&dl=ACM&coll=portal&CFID=11111111&CFTOKEN=2222222 - [Similar pages](#)[[More results from portal.acm.org](#)]**1998 Master's Thesis (Jun-ichi TAKEUCHI)**


A Method for Designing a Multi-monitoring System of the Dismantling

Tele-operation System. The Dismantling Tele-operation System ...

kanri.mech.okayama-u.ac.jp/1998/atakeuchi.htm - 3k - [Cached](#) - [Similar pages](#)**[PDF] DR. JUNICHI TAKEUCHI PAST INTERNATIONAL DIRECTOR**File Format: PDF/Adobe Acrobat - [View as HTML](#)**DR. JUNICHI TAKEUCHI PAST INTERNATIONAL DIRECTOR** Dr. Junichi "Jun" Takeuchi,

of Handa, Japan, was elected to serve a two-year term as a director of The ...

www.lionsclubs.org/EN/content/pastdir/takeuchi.pdf - [Similar pages](#)


[Web](#) [Images](#) [Groups](#) [News](#) [Froogle](#)^{New!} [more »](#)

[Advanced Search](#)
[Preferences](#)

WebResults 1 - 10 of about 65 for "**smart sifter**" OR "**smartsifter**". (0.37 seconds)Did you mean: "smart sifter" OR "**smartsoft**"**SmartSifter** 外れ値検出エンジン

現在、**SmartSifter**はログ解析ツールGateMinderに
コアエンジンとして組み込まれております。...

www.labs.nec.co.jp/DTmining/products/smartsifter/ - 14k - [Cached](#) - [Similar pages](#)
SmartSifter - 製品の概要

... 計算時間が多大である、オンラインでリアルタイムに処理できない等といった問題を抱えていました。 **SmartSifter** は ...

www.labs.nec.co.jp/DTmining/products/s_sifter/ - 10k - [Cached](#) - [Similar pages](#)
[[More results from www.labs.nec.co.jp](#)]On-Line Unsupervised Outlier Detection Using Finite Mixtures with ...

... **SmartSifter** is an outlier detection engine addressing this problem from the viewpoint of statistical learning theory. This paper ...

portal.acm.org/citation.cfm?id=982721.982727&dl=portal&dl=GUIDE&CFID=11111111&CFTOKEN=2222222 - [Similar pages](#)

Discovering outlier filtering rules from unlabeled data

... In prior work we have developed **SmartSifter**, which is an on-line outlier detection algorithm based on unsupervised learning from data. ...

portal.acm.org/citation.cfm?id=502570&dl=ACM&coll=portal&CFID=11111111&CFTOKEN=2222222 - [Similar pages](#)

[id=502570&dl=ACM&coll=portal&CFID=11111111&CFTOKEN=2222222](http://portal.acm.org/citation.cfm?id=502570&dl=ACM&coll=portal&CFID=11111111&CFTOKEN=2222222) - [Similar pages](#)

[[More results from portal.acm.org](#)][PS] On-line Unsupervised Outlier Detection Using FiniteMixtures with ...

File Format: Adobe PostScript - [View as Text](#)

... **SmartSifter**, which we abbreviate as SS, is an outlier detection engine addressing this problem from the viewpoint of statistical learning theory. ...

research.cmis.csiro.au/gjw/papers/kdd01.ps - [Similar pages](#)

[PDF] Evolving Fuzzy Classifiers for Intrusion Detection

File Format: PDF/Adobe Acrobat - [View as HTML](#)

... is shown in table 4. Algorithm FA % DR % Complexity EFRID 7.0 98.95 O(n) RIPPER-Artificial Anomalies [20] 2.02 94.26 O(n*log 2 n) **SMARTSIFTER** [21] - 82.0 O(n 2 ...

seclab1.cs.memphis.edu/people/jgomez/papers/infassu2002.pdf - [Similar pages](#)

[PDF] Mining the Data Stream

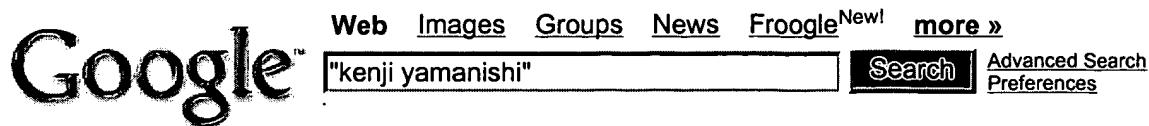
File Format: PDF/Adobe Acrobat - [View as HTML](#)

Page 1. Data Streams Data Mining Moving Forward Mining the Data Stream

Graham Williams CSIRO Data Mining GPO Box 664, Canberra 2601 ...

datamining.csiro.au/papers/his03.pdf - [Similar pages](#)

2003 Equipment Catalog - Buyer's Guide - from GRAINNET

**Web**

Results 1 - 10 of about 741 for "kenji yamanishi". (0.27 seconds)

Kenji Yamanishi's resume

Kenji Yamanishi. Last Updated: April. 2004. Japanese Version. Research Fellow.

Data Mining Research Group,. Ineternet Systems Laboratories, NEC Corporation. ...

www.labs.nec.co.jp/DTmining/members/ yamanishi/index_e.html - 101k - Apr 29, 2004 - [Cached](#) - [Similar pages](#)**Kenji Yamanishi's resume**山西 健司 (**Kenji Yamanishi**) . Last Updated: April 2004. English

version. 所属. NEC インターネットシステム研究所 ...

www.labs.nec.co.jp/DTmining/members/ yamanishi/index_j.html - 101k - Apr 29, 2004 - [Cached](#) - [Similar pages](#)[[More results from www.labs.nec.co.jp](#)]**DBLP: Kenji Yamanishi**dblp.uni-trier.de **Kenji Yamanishi**. ... 2003. 30, EE, Satoshi Morinaga, **Kenji Yamanishi**,

Jun-ichi Takeuchi: Distributed cooperative mining for information consortia. ...

www.informatik.uni-trier.de/~ley/db/indices/ a-tree/y/Yamanishi:Kenji.html - 15k - Apr 28, 2004 - [Cached](#) - [Similar pages](#)**DBLP: Hang Li**... 18, EE, Hang Li, **Kenji Yamanishi**: Topic analysis using a finite mixture model. Inf. ... 14,EE, **Kenji Yamanishi**, Hang Li: Mining Open Answers in Questionnaire Data. ...www.informatik.uni-trier.de/~ley/ db/indices/a-tree/l/Li:Hang.html - 15k - Apr 28, 2004 - [Cached](#) - [Similar pages](#)[[More results from www.informatik.uni-trier.de](#)]**Kenji Yamanishi****Kenji Yamanishi**. **Kenji Yamanishi**. A loss bound model for on-line stochastic prediction algorithms. ... Abstract, References, and Citations. **Kenji Yamanishi**. ...theory.lcs.mit.edu/~iandc/Authors/yamanishikenji.html - 2k - [Cached](#) - [Similar pages](#)**Mining product reputations on the Web**... 216-8555, JAPAN. **Kenji Yamanishi**, NEC Corporation, 4-1-1, Miyazaki,

Miyamae, Kawasaki, Kanagawa, 216-8555, JAPAN. Kenji Tateishi, NEC ...

portal.acm.org/ citation.cfm?id=775098&dl=ACM&coll=GUIDE&CFID=11111111&CFTOKEN=2222222 -

[Similar pages](#)**Text classification using ESC-based stochastic decision lists**... **Kenji Yamanishi**, NEC Corporation, 4-1-1 Miyazaki Miyamakc-ku, Kawasaki, 216-8555 Japan. Publisher, Pergamon Press, Inc. Tarrytown, NY, USA. ...

portal.acm.org/ citation.cfm?

id=638647&dl=ACM&coll=GUIDE&CFID=11111111&CFTOKEN=2222222 - [Similar pages](#)[[More results from portal.acm.org](#)]**kenji yamanishi - ResearchIndex document query**Searching for PHRASE **kenji yamanishi**. Restrict to: Header Title Order by: Expected

citations Hubs Usage Date 11 documents found. Order: number of citations. ...

citeseer.ist.psu.edu/cis?q=Kenji+Yamanishi - 11k - [Cached](#) - [Similar pages](#)


[Web](#) [Images](#) [Groups](#) [News](#) [Froogle](#)^{New!} [more »](#)

[Advanced Search](#)
[Preferences](#)

Web

 Results 1 - 10 of about 3,520 for "[outlier detection](#)" "[data mining](#)". (0.55 seconds)

[PDF] [Algorithms for spatial outlier detection - Data Mining, 2003. ICDM ...](#)

 File Format: PDF/Adobe Acrobat - [View as HTML](#)

... the experimental results for all four spa- tial **outlier detection** algorithms ... the Proceedings of the Third IEEE International Conference on **Data Mining** (ICDM'03 ...

europa.nvc.cs.vt.edu/~ctl/Publication/ICDM-2003-P.pdf - [Similar pages](#)

Sponsored Links

[Data Mining](#)

Buy The Latest Releases & Popular Titles at Barnes & Noble. Shop Now www.BarnesandNoble.com

[See your message here...](#)

[KDD-2001 tutorial: Data Mining for Outliers with Robust Statistics](#)

... outliers. Indeed, the topic of **outlier detection** has received relatively little attention in the **data mining** and KDD literature. Yet ...

www.acm.org/sigkdd/kdd2001/Tutorials/kdd2001_T2.html - 6k - Apr 29, 2004 - [Cached](#) - [Similar pages](#)

[Outlier Detection for High Dimensional Data - Aggarwal, Yu ...](#)

... Aggarwal (2001) (Correct) 0.4: Survey Of Clustering **Data Mining** Techniques - Berkhin ... Aggarwal, Yu (2001) (Correct) 0.6: **Outlier Detection** Integrating Semantic ... citeseer.ist.psu.edu/aggarwal01outlier.html - 23k - [Cached](#) - [Similar pages](#)

[PPT] [Data Mining for Network Intrusion Detection](#)

 File Format: Microsoft Powerpoint 97 - [View as HTML](#)

... Learning from Rare Classes. **Data mining** technique for misuse detection. ... Generalization of over-sampling technique. Anomaly Detection – **Outlier Detection** Schemes. ...

www.eecs.tulane.edu/Peng/cpsc464/BethIDDMpresentation.ppt - [Similar pages](#)

[PPT] [Data Mining in Cyber Threat Analysis](#)

 File Format: Microsoft Powerpoint 97 - [View as HTML](#)

... Anomaly/**outlier detection**. Summarization of attacks using association pattern analysis. ... Standard **data mining** models are not suitable for rare classes. ...

www-users.cs.umn.edu/~kumar/Presentation/minds.ppt - [Similar pages](#)

[PDF] [A Comparative Study of RNN for Outlier Detection in Data Mining](#)

 File Format: PDF/Adobe Acrobat - [View as HTML](#)

A Comparative Study of RNN for **Outlier Detection** in **Data Mining** Graham Williams Rohan Baxter Hongxing He Simon Hawkins Lifang Gu CSIRO Enterprise **Data Mining** ... www.act.cmis.csiro.au/rohanb/PAPERS/icdm02.pdf - [Similar pages](#)

[PDF] [OUTLAW: Using Geo-Spatial Associations for Outlier Detection and ...](#)

 File Format: PDF/Adobe Acrobat - [View as HTML](#)

... OUTLAW, which is currently under development, enables **outlier detection** and visual analysis using **data mining** techniques that rely on geo-spatial associations. ...

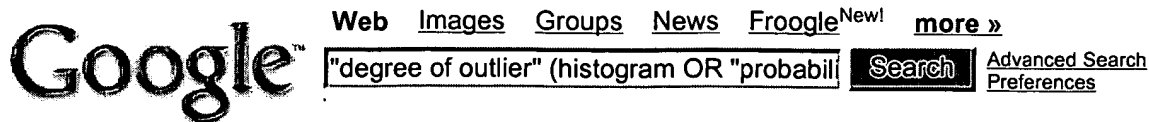
cimic.rutgers.edu/~vandana/pubs/outlaw_cr.pdf - [Similar pages](#)

[PPT] [Spatial Data Mining ---- Classification and Clustering](#)

 File Format: Microsoft Powerpoint 97 - [View as HTML](#)

Spatial **Data Mining** --- Classification and **Outlier Detection**. Presented by Shengnan Cong. Presented by Shengnan Cong. 2. 3/9/04. Outline. ...

www-courses.cs.uiuc.edu/~cs412/slides/Spatial3Cong.ppt - [Similar pages](#)



Web Results 1 - 9 of about 12 for "**degree of outlier**" (histogram OR "probability density"). (0.22 seconds)

[PDF] Iris Recognition: An Emerging Biometric Technology - Proceedings ...

File Format: PDF/Adobe Acrobat - [View as HTML](#)

... The **histogram**-based approach to model fitting should avoid problems with local minima that the active contour model's gradient descent procedure might ...

www.cs.yorku.ca/~wildes/wildesPIEEE1997.pdf - [Similar pages](#)

[PDF] Robust projection indices

File Format: PDF/Adobe Acrobat - [View as HTML](#)

... The **degree of outlier** contamination in the density is specified by $\frac{1}{4} - \frac{1}{2}$ (although, of course, for large n the outlier is no longer outlying). ...

www.stats.bris.ac.uk/~guy/Research/papers/tindex.pdf - [Similar pages](#)

[PDF] Robust projection indices

File Format: PDF/Adobe Acrobat - [View as HTML](#)

... outlier centred on $\frac{1}{4} \mu$. The **degree of outlier** contamination in the density is specified by $\frac{1}{4} - \frac{1}{2}$. The marginal density ...

www.stats.bris.ac.uk/pub/ResRept/2000/05.pdf - [Similar pages](#)

[[More results from www.stats.bris.ac.uk](#)]

[PDF] Survey of Clustering Data Mining Techniques

File Format: PDF/Adobe Acrobat - [View as HTML](#)

... Describing the numbers of data points per every unit represents an extreme case of clustering, a **histogram**, where no actual clustering takes place. ...

www.accrue.com/products/rp_cluster_review.pdf - [Similar pages](#)

[PS] Motion-based video segmentation using continuation method

File Format: Adobe PostScript - [View as Text](#)

... To date numerous methods have been proposed for segmentation of still images; **histogram**-based, clusteringbased, region ... control the **degree of outlier** rejection. ...

www2.inrs-telecom.quebec.ca/users/viscom/francais/recherche/local/cpapers/Yang98vcip.ps.gz - [Similar pages](#)

[PDF] FindOut: Finding Outliers in Very Large Datasets

File Format: PDF/Adobe Acrobat

... Recently, Breunig et al. (2000) proposed the notion of LOF (local outlier factor) to indicate the **degree of outlier**-ness for each data point in a dataset. ...

www.springerlink.com/index/NTM59P9DB1AADH03.pdf - [Similar pages](#)

[PDF] The Patents & Designs Journal 5817

File Format: PDF/Adobe Acrobat - [View as HTML](#)

Page 1. An Executive Agency of the Department of Trade and Industry
The Patents & Designs Journal 8 November 2000 No.5817 Application ...

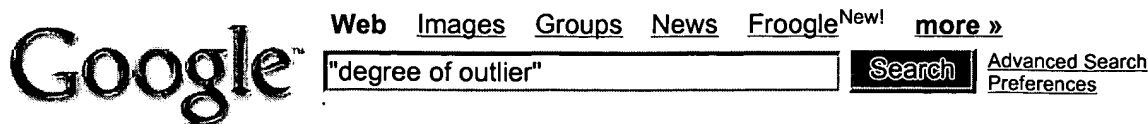
www.patent.gov.uk/patent/notices/journals/2000/5817.pdf - [Similar pages](#)

[PDF] Robust projection indices

File Format: PDF/Adobe Acrobat

... $Y \square x, y \square$ models two large clusters on the horizontal axis separated by 2 with a movable outlier centred on $\square 0, \square$. The **degree of outlier** contamination in ...

www.blackwell-synergy.com/links/doi/10.1111/1467-9868.00298/abs/ - Supplemental Result - [Similar pages](#)

**Web**Results 1 - 10 of about 35 for "**degree of outlier**". (0.09 seconds)[#include <stdlib.h> #include <math.h> #include <vector> #include ...](#)

Use this multiplier to convert // floating-point weights to integers. static

const int multiplier = 100; // The **degree of outlier** connections. ...www.cs.berkeley.edu/projects/vision/grouping/segbench/code/Benchmark/match.cc - 13k - [Cached](#) -[Similar pages](#)[Journal of the Royal Statistical Society: Series B \(Statistical ...](#)

... properties of one new Student t-index and Hall's index and show that the new index

automatically adapts its robustness to the **degree of outlier** contamination ...www.blackwellpublishing.com/abstract.asp?ref=1369-7412&vid=63&iid=3&aid=298&s=&site=1 - 14k - [Cached](#) -[Similar pages](#)[\[PDF\] Outlier Detection Under Interval and Fuzzy Uncertainty ...](#)File Format: PDF/Adobe Acrobat - [View as HTML](#)... We also provide algorithms that estimate the **degree of "outlier-ness"** of a

given value – measured as the largest value for which is outside the ...

www.cs.utep.edu/vladik/2003/tr03-10d.pdf - [Similar pages](#)[Robust projection indices](#)

... properties of one new Student's t index and Hall's index and show that the new index

automatically adapts its robustness to the **degree of outlier** contamination ...www.stats.bris.ac.uk/pub/ResRept/2000/05.html - 2k - [Cached](#) - [Similar pages](#)[\[doc\] GENERALIZATION TO HIGHER DIMENSIONS](#)File Format: Microsoft Word 2000 - [View as HTML](#)... nearest neighbor of p. It ranks **degree of outlier-ness**; k is the kth

nearest neighbor; n is the number of □top□ outliers you want; ...

www.cs.rpi.edu/~zaki/dmcourse/notes/11-13-03/11-13-03.doc - [Similar pages](#)[\[PS\] \[ProductName: GNU Ghostscript \]](#)File Format: Adobe PostScript - [View as Text](#)

... Another method is the Dkn outlier method, where Dk(p) is the distance to

the k-th nearest neighbor of p. * It ranks **degree of outlier-ness** *. ...www.cs.rpi.edu/~zaki/dmcourse/notes/11-13-03/11-13-03.ps - [Similar pages](#)[\[More results from www.cs.rpi.edu \]](#)[Regression Diagnostics - Outliers](#)... to its own fitted value; Measure of **degree of outlier** in predictor;

Useful for determining influence of each observation. Note that:

www.geog.ucsb.edu/~joel/g210_w04/lecture_notes/lect13/oh04_13_1.html - 5k - [Cached](#) - [Similar pages](#)[\[PDF\] DRAFT DRAFT DRAFT](#)File Format: PDF/Adobe Acrobat - [View as HTML](#)... 6.2 How will outliers be identified? What **degree of outlier** status will

evoke a corrective action plan or a quality improvement initiative? ...

www.dhhs.state.nc.us/dph/taskforce/docs/ACCOUNT_COMM_OVERVIEW_9-29-03.pdf - [Similar pages](#)[\[PDF\] Random Voronoi Ensembles for gene selection](#)File Format: PDF/Adobe Acrobat - [View as HTML](#)

Jun-ichi Takeuchi, Dr.Eng.

Principal Researcher in Data Mining Research Group

Internet Systems Research Laboratories, NEC Corporation

1753, Shimonumabe, Nakahara-ku, Kawasaki, Kanagawa, 211-8666, Japan.

E-mail: tak@ap.jp.nec.com

Japanese version

- Research (in Japanese)
 - Data mining.
 - Universal noiseless coding and universal prediction.
 - Information geometry of Bayesian statistics
 - Computational learning theory.
- Publication list
- Professional Activities
- Curriculum Vitae
- Biography

Last updated: April 21st, 2004.

[Go to the top of NEC R&D on Data & Text Mining](#)

Publication List

Refereed Journal Papers

1. K. Yamanishi, J. Takeuchi, G. Williams, and P. Milne:
"On-line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms," *Data Mining and Knowledge Discovery Journal*, 8 (3): 275-300, May 2004.
2. N. Abe, J. Takeuchi, and M. Warmuth :
"Polynomial Learnability of Stochastic Rules with respect to the KL-divergence and Quadratic Distance," *IEICE trans. (D)*, Vol.E84-D No.3 pp. 299-316, 2001.
3. J. Takeuchi, N. Abe, and S. Amari :
"The Lob-Pass problem," *Journal of Computer and System Sciences*, Vol. 61, No. 3, pp. 523-557, 2000.
4. A. Nakamura, J. Takeuchi, and N. Abe :
"Efficient distribution-free population learning of simple concepts," *Annals of Mathematics and Artificial Intelligence*, 23, pp. 53-82, 1998.
5. J. Takeuchi :
"Characterization of the Bayes estimator and the MDL estimator for exponential families," *IEEE trans. Information Theory*, Vol. 43, No. 4, pp. 1165-1174, 1997.
6. J. Takeuchi :
"Improved sample complexity bounds for parameter estimation," *IEICE trans. (D)*, Vol. E78D, No. 5, pp. 526-531, 1995.

Refereed Conference Papers

1. S. Morinaga, K. Yamanishi, & J. Takeuchi:
"Distributed Cooperative Mining for Information Consortia", *Proc. of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press (KDD2003)*, 2003.
2. J. Takeuchi & K. Yamanishi :
"A Unifying Framework for Detecting Outliers and Change Points from Non-Stationary Time Series (in Japanese)", in *Proc. of the fifth Workshop on Information-Based Induction Sciences (IBIS2002)*, 2002.
3. K. Yamanishi & J. Takeuchi:
"A Unifying Framework for Detecting Outliers and Change Points from Non-Stationary Time Series Data", *Proc. of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press (KDD2002)*, 2002.
4. J. Takeuchi, T. Kawabata, and A. R. Barron:
"Properties of Jeffreys mixture for Markov sources", in *Proc. of the fourth Workshop on Information-Based Induction Sciences (IBIS2001)*, pp. 327-332, 2001.
5. K. Yamanishi & J. Takeuchi:

- ``Discovering Outlier Filtering Rules from Unlabeled Data --Combining a Supervised Learner with an Unsupervised Learner--'' (in Japanese)", in *Proc. the fourth Workshop on Information-Based Induction Sciences (IBIS2001)*, pp. 111-116, 2001.
6. K. Yamanishi & J. Takeuchi:
``Discovering Outlier Filtering Rules from Unlabeled Data --Combining Supervised Learners with Unsupervised Learners--'', in *Proc. of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, (KDD2001)*, 2001.
 7. K. Yamanishi and J. Takeuchi:
``Statistical Outlier Detection Using On-line Discounting Learning Algorithms'' (in Japanese), in *Proc. the third Workshop on Information-Based Induction Sciences*, 2000.
 8. J. Takeuchi :
``On minimax regret with respect to families of stationary stochastic processes'' (in Japanese), in *Proc. of the third Workshop on Information Based Induction Sciences*, 2000.
 9. K. Yamanishi, J. Takeuchi, G. Williams, and P. Milne:
``On-line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms'', in *Proc. of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press (KDD2000)*, pp:320-324, 2000.
 10. J. Takeuchi & A. R. Barron :
``Asymptotically minimax regret by Bayes mixtures'', in *Proc. of 1998 IEEE International Symposium on Information Theory*, 1998.
 11. J. Takeuchi & T. Kawabata :
``Approximation of Bayes code for Markov sources'', in *Proc. of 1995 IEEE International Symposium on Information Theory*, p.391, 1995.
 12. J. Takeuchi :
``Characterization of the Bayes estimator and the MDL estimator for exponential families'', in *Proc. of 1995 IEEE International Symposium on Information Theory* (long presentation), p.228, 1995.
 13. A. Nakamura, N. Abe & J. Takeuchi :
``Efficient distribution-free population learning of simple concepts'', in *Proc. of the 5th International Workshop on Algorithmic Learning Theory*, pp. 500-515, 1994.
 14. N. Abe & J. Takeuchi :
``The 'Lob-Pass' problem and an on-line learning model of rational choice'', in *Proc. of the 7th Annual Conference on Computational Learning Theory*, pp. 422-428, 1993.
 15. J. Takeuchi :
``Some improved sample complexity bounds in the probabilistic PAC learning model'', in *Proc. of the 3rd Workshop on Algorithmic Learning Theory*, pp. 208-219, 1992.
 16. N. Abe, J. Takeuchi, and M. Warmuth :
``Polynomial learnability of probabilistic concepts with respect to the Kullback-Leibler divergence'', in *Proc. of the 4th annual Workshop on Computational Learning Theory*, pp. 277-289, 1991.

Articles and other Invited Papers

1. N. Abe, K. Yamanishi, A. Nakamura, H. Mamitsuka, J. Takeuchi, & H. Li:
``Distributed and Active Learning," The Foundations of Real-World Intelligence, Oct. 2001.
2. J. Takeuchi & K. Yamanishi:
``Statistical outlier detection in data mining" (in Japanese), *Bulletin of the Japan Society for Industrial and Applied Mathematics*, Vol. 10, No. 3, 2001.
3. J. Takeuchi:
``Asymptotically Minimax Codes by Bayes Procedures" (in Japanese), in *Proc. of IEICE Society Conference*, October 1998.
4. J. Takeuchi:
``Stochastic complexity and Jeffreys mixture prediction strategies" (in Japanese), in *Proc. of the first Workshop on Information Based Induction Sciences*, pp. 9-16, 1998.
5. A. R. Barron & J. Takeuchi:
``Mixture models achieving optimal coding regret," in *Proc. of 1998 IEEE Inform. Theory Workshop*, 1998.

Other Conference Papers (selected)

1. J. Takeuchi & A. R. Barron:
``Robustly minimax codes for universal data compression", in *Proc. of the 21st Symposium on Information Theory and its Applications (SITA'98)*, 1998.
2. J. Takeuchi & A. R. Barron:
``Asymptotically minimax regret for exponential families", in *Proc. of the 20th Symposium on Information Theory and its Applications (SITA'97)*, pp. 665-668, 1997.
Best papers award at SITA'97.
3. J. Takeuchi & S. Amari:
``The alpha-parallel prior and its properties" (in Japanese), in *Technical Report of IEICE*, IT26-20, pp. 61-66, 1996.
4. J. Takeuchi & K. Kawabata:
``On data compression algorithms by Bayes coding for Markov sources" (in Japanese), in *Proc. of the 17th Symposium on Information Theory and its Applications (SITA'94)*, pp. 513-516, 1994.

April 21st, 2004

Go to [the top of Jun'ichi Takeuchi's page](#)

Go to [the top of NEC R&D on Data & Text Mining](#)

Jun-ichi Takeuchi

List of publications from the [DBLP Bibliography Server](#) - [FAQ](#)

[Coauthor Index](#) - Ask others: [ACM DL](#) - [ACM Guide](#) - [CiteSeer](#) - [CSB](#) - [Google](#)

2003		
10	EE	Satoshi Morinaga , Kenji Yamanishi , Jun-ichi Takeuchi: Distributed cooperative mining for information consortia. KDD 2003 : 619-624
2002		
9	EE	Kenji Yamanishi , Jun-ichi Takeuchi: A unifying framework for detecting outliers and change points from non-stationary time series data. KDD 2002 : 676-681
2001		
8	EE	Kenji Yamanishi , Jun-ichi Takeuchi: Discovering outlier filtering rules from unlabeled data: combining a supervised learner with an unsupervised learner. KDD 2001 : 389-394
2000		
7	EE	Kenji Yamanishi , Jun-ichi Takeuchi, Graham J. Williams , Peter Milne : On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. KDD 2000 : 320-324
6		Jun-ichi Takeuchi, Naoki Abe , Shun-ichi Amari : The Lob-Pass Problem. J. Comput. Syst. Sci. 61(3): 523-557 (2000)
1998		
5		Atsuyoshi Nakamura , Jun-ichi Takeuchi, Naoki Abe : Efficient Distribution-Free Population Learning of Simple Concepts. Ann. Math. Artif. Intell. 23(1-2): 53-82 (1998)
1994		
4		Atsuyoshi Nakamura , Naoki Abe , Jun-ichi Takeuchi: Efficient Distribution-free Population Learning of Simple Concepts. AII/ALT 1994 : 500-515
1993		
3	EE	Naoki Abe , Jun-ichi Takeuchi: The "lob-pass" Problem and an On-line Learning Model of Rational Choice. COLT 1993 : 422-428
1992		
2		Jun-ichi Takeuchi: Some Improved Sample Complexity Bounds in the Probabilistic PAC Learning Model. ALT 1992 : 208-219
1991		
1	EE	Naoki Abe , Manfred K. Warmuth , Jun-ichi Takeuchi: Polynomial Learnability of Probabilistic Concepts with Respect to the Kullback-Leibler Divergence. COLT 1991 : 277-289

Coauthor Index

1	Naoki Abe	[1] [3] [4] [5] [6]
2	Shun-ichi Amari	[6]
3	Peter Milne	[7]
4	Satoshi Morinaga	[10]
5	Atsuyoshi Nakamura	[4] [5]
6	Manfred K. Warmuth	[1]
7	Graham J. Williams	[7]
8	Kenji Yamanishi	[7] [8] [9] [10]

DBLP: [[Home](#) | Search: [Author](#), [Title](#) | [Conferences](#) | [Journals](#)]

Michael Ley (ley@uni-trier.de) Fri Apr 30 16:27:21 2004

Kenji Yamanishi

List of publications from the [DBLP Bibliography Server](#) - [FAQ](#)

[Coauthor Index](#) - Ask others: [ACM DL](#) - [ACM Guide](#) - [CiteSeer](#) - [CSB](#) - [Google](#)

2003		
30	EE	Satoshi Morinaga , Kenji Yamanishi , Jun-ichi Takeuchi : Distributed cooperative mining for information consortia. KDD 2003 : 619-624
29	EE	Hang Li , Kenji Yamanishi : Topic analysis using a finite mixture model. Inf. Process. Manage. 39(4): 521-541 (2003)
2002		
28	EE	Satoshi Morinaga , Kenji Yamanishi , Kenji Tateishi , Toshikazu Fukushima : Mining product reputations on the Web. KDD 2002 : 341-349
27	EE	Kenji Yamanishi , Jun-ichi Takeuchi : A unifying framework for detecting outliers and change points from non-stationary time series data. KDD 2002 : 676-681
26	EE	Kenji Yamanishi , Hang Li : Mining Open Answers in Questionnaire Data. IEEE Intelligent Systems 17(5): 58-63 (2002)
25		Hang Li , Kenji Yamanishi : Text classification using ESC-based stochastic decision lists. Inf. Process. Manage. 38(3): 343-361 (2002)
2001		
24	EE	Kenji Yamanishi , Jun-ichi Takeuchi : Discovering outlier filtering rules from unlabeled data: combining a supervised learner with an unsupervised learner. KDD 2001 : 389-394
23	EE	Hang Li , Kenji Yamanishi : Mining from open answers in questionnaire data. KDD 2001 : 443-449
2000		
22	EE	Kenji Yamanishi , Jun-ichi Takeuchi , Graham J. Williams , Peter Milne : On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms. KDD 2000 : 320-324
1999		
21	EE	Kenji Yamanishi : Extended Stochastic Complexity and Minimax Relative Loss Analysis. ATL 1999 : 26-38
20	EE	Hang Li , Kenji Yamanishi : Text Classification Using ESC-based Stochastic Decision Lists. CIKM 1999 : 122-130
19		Kenji Yamanishi : Distributed Cooperative Bayesian Learning Strategies. Inf. Comput. 150(1): 22-56 (1999)
1998		
18	EE	Kenji Yamanishi : Minimax Relative Loss Analysis for Sequential Prediction Algorithms Using

		Parametric Hypotheses. <u>COLT</u> 1998: 32-43
1997		
17		<u>Hang Li</u> , Kenji Yamanishi: Document Classification Using a Finite Mixture Model. <u>ACL</u> 1997: 39-47
16	EE	Kenji Yamanishi: Distributed Cooperative Bayesian Learning Strategies. <u>COLT</u> 1997: 250-262
15		Kenji Yamanishi: On-Line Maximum Likelihood Prediction with Respect to General Loss Functions. <u>J. Comput. Syst. Sci.</u> 55(1): 105-118 (1997)
1996		
14	EE	Kenji Yamanishi: A Randomized Approximation of the MDL for Stochastic Models with Hidden Variables. <u>COLT</u> 1996: 99-109
1995		
13	EE	Kenji Yamanishi: Randomized Approximate Aggregating Strategies and Their Applications to Prediction and Discrimination. <u>COLT</u> 1995: 83-90
12		Kenji Yamanishi: On-line maximum likelihood prediction with respect to general loss functions. <u>EuroCOLT</u> 1995: 84-98
11		<u>Hiroshi Mamitsuka</u> , Kenji Yamanishi: alpha-Helix region prediction with stochastic rule learning. <u>Computer Applications in the Biosciences</u> 11(4): 399-411 (1995)
10		Kenji Yamanishi: A Loss Bound Model for On-Line Stochastic Prediction Algorithms <u>Inf. Comput.</u> 119(1): 39-54 (1995)
9		Kenji Yamanishi: Probably Almost Discriminative Learning. <u>Machine Learning</u> 18(1): 23-50 (1995)
1994		
8	EE	Kenji Yamanishi: The Minimum L -Complexity Algorithm and its Applications to Learning Non-Parametric Rules. <u>COLT</u> 1994: 173-182
1993		
7	EE	Kenji Yamanishi: On Polynomial-Time Probably almost Discriminative Learnability. <u>COLT</u> 1993: 94-100
1992		
6		<u>Hiroshi Mamitsuka</u> , Kenji Yamanishi: Protein Secondary Structure Prediction Based on Stochastic-Rule Learning. <u>ALT</u> 1992: 240-251
5	EE	Kenji Yamanishi: Probably Almost Discriminative Learning. <u>COLT</u> 1992: 164-171
4		Kenji Yamanishi: A Learning Criterion for Stochastic Rules. <u>Machine Learning</u> 9: 165-203 (1992)
1991		
3	EE	Kenji Yamanishi: A Loss Bound Model for On-Line Stochastic Prediction Strategies. <u>COLT</u> 1991: 290-302
2		Kenji Yamanishi, <u>Akihiko Konagaya</u> : Learning Stochastic Motifs from Genetic Sequences. <u>ML</u> 1991: 467-471
1990		
1	EE	Kenji Yamanishi: A Learning Criterion for Stochastic Rules. <u>COLT</u> 1990: 67-81

Coauthor Index

1	<u>Toshikazu Fukushima</u>	[28]
2	<u>Akihiko Konagaya</u>	[2]
3	<u>Hang Li</u>	[17] [20] [23] [25] [26] [29]
4	<u>Hiroshi Mamitsuka</u>	[6] [11]
5	<u>Peter Milne</u>	[22]
6	<u>Satoshi Morinaga</u>	[28] [30]
7	<u>Jun-ichi Takeuchi</u>	[22] [24] [27] [30]
8	<u>Kenji Tateishi</u>	[28]
9	<u>Graham J. Williams</u>	[22]

DBLP: [[Home](#) | [Search: Author, Title](#) | [Conferences](#) | [Journals](#)]
Michael Ley (ley@uni-trier.de) Fri Apr 30 16:27:21 2004



Graham Williams: Publications



Last Modified 2000/05/15 14:03:34 Graham.Williams@cmis.csiro.au

Publications on the following topics are included:

- Data Mining and Machine Learning
 - Spatial Reasoning
 - Databases and Legacy Systems
 - Knowledge Representation
-

Data Mining and Machine Learning

Outlier Detection Using Replicator Neural Networks

Hongxing He, Simon Hawkins, Graham Williams, Rohan Baxter
DaWaK 2002

Mining Temporal Patterns from Health Care Data

Weiqiang Lin, Mehmet Orgun, Graham Williams
DaWaK 2002

Feature Selection for Pathology Laboratory Monitoring

Simon Hawkins, Graham Williams, Rohan Baxter
Topics in Health Information Management, Volume 22 Number 1 August 2001 Pages 14-23

The Pathology Explorer

Graham Williams and Simon Hawkins
CMIS Technical Report Number 01/116
Report in Confidence to the Health Insurance Commission

Advances in Knowledge Discovery and Data Mining

David Cheung, Graham J. Williams, Qing Li
5th Pacific Asia Conference , PAKDD 2001, Hong Kong, China, April 2001, Proceedings.
Lecture Notes in Artificial Intelligence, Volume 2035, Springer.

ISBN 3540419101

Feature Selection for Temporal Health Records

Rohan A. Baxter, Graham J. Williams, Hongxing He
in *Advances in Knowledge Discovery and Data Mining*,
Edited by David Cheung, Graham J. Williams, Qing Li
Lecture Notes in Artificial Intelligence, Volume 2035, Springer, April 2001.
Proceedings of the 5th Pacific Asia Conference on Knowledge Discovery and Data Mining
PAKDD 2001, Hong Kong, China.

Temporal Data Mining Using Hidden Markov-Local Polynomial Models

Weiqian Lin, Mehmet A. Orgun, Graham J. Williams
in *Advances in Knowledge Discovery and Data Mining*,
Edited by David Cheung, Graham J. Williams, Qing Li
Lecture Notes in Artificial Intelligence, Volume 2035, Springer, April 2001.
Proceedings of the 5th Pacific Asia Conference on Knowledge Discovery and Data Mining
PAKDD 2001, Hong Kong, China.

Data Mining of Administrative Claims Data for Pathology Services

Simon Hawkins, Graham Williams, Rohan Baxter, Peter Christen, Michael Fett, Markus Hegland,
Fuchun Huang, Ole Nielsen, Tatiana Semanova, Andrew Smith
Hawaii International Conference on System Sciences (HICSS-35), Data Mining in Health,
January 2001, Hawaii, USA

Temporal data mining using multi-level local polynomial models

Weiqiang Lin, Mehmet A. Orgun, Grhama J. Williams
In *Second International Conference on Intelligent Data Engineering and Automated Learning*
Lecture Notes in Computer Science, Springer, December 2000.
IDEAL 2000, Hong Kong.

On-line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms

[ps]

Kenji Yamanishi, Jun-ichi Takeuchi, Graham Williams, Peter Milne
Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining
KDD-01, August 20-23, 2000, Boston, MA USA

Applications of Artificial Intelligence in Industry

Dickson Lukose, Graham Williams (editors)
Proceedings of the Symposium on the Application of Artificial Intelligence in Industry.
Melbourne, Australia, August 2000
ISBN 0730027937

Mining Taxation Data with Parallel BMARS[[pdf](#) [ps](#)]

Sergey Bakin, Markus Hegland, and Graham Williams

Parallel Algorithms and Applications

Vol 15, pp 37-55, May 2000

The Integrated Delivery of Large-Scale Data Mining: The ACSys Data Mining Project

Graham Williams, Irfan Altas, Sergey Bakin, Peter Christen, Markus Hegland,

Alonso Marquez, Peter Milne, Rajehndra Nagappan, and Stephen Roberts

In Large-Scale Parallel Data Mining, State-of-the-Art Survey

Edited by Mohammed J. Zaki and Ching-Tien Ho

Lecture Notes in Artificial Intelligence, Volume 1759

Springer-Verlag, 2000

Data Mining Tools

Irfan Altas, Sergey Bakin, Markus Hegland, Stephen Roberts, Berwin Turlach, and Graham Williams

IEEE Transactions on Concurrency

Submitted 1999

An Overview of ACSys Data Mining

Graham J. Williams

Computational Techniques and Applications Conference and Workshops
(CTAC99)

Canberra, September 1999

Integrated Delivery of Data Mining

Graham J. Williams

KDD'99 Workshop on Large-Scale Parallel KDD Systems

San Diego, August 1999

Evolving Interestingness for Data Mining

Graham J. Williams

Third Pacific-Asia Conference on Knowledge Discovery and Data Mining

Beijing, April 1999

Data Mining Tutorial

Graham J. Williams

SEAL'98

Canberra, November 1998

Evolvolutionary Techniques in Data Mining Interestingness

Graham J. Williams
Workshop on Evolutionary Computation
Canberra, October 1998

The Data Miner's Arcade: Pluggable Data Mining

Graham J. Williams
Technical Report
May 1998.

Abstract

The Data Miner's Arcade is a Java-based environment for data mining. It implements an Object-Oriented model for the Data Mining process, with **standard** interfaces for accessing data and for delivering results. By developing standards, new tools can plug into the environment with a minimum of effort, providing 'Plug-n-Play' opportunities with new tools as they become available. Data can be accessed from Database systems through ODBC and JDBC, or from other sources and managed internally within the Arcade. The Extensible Markup Language (XML) is used as the target "language" for all Data Mining tools within the environment. The Predictive Modelling Markup Language (PMML) developed by UIC is an example of the XML markup that the system handles. Data Mining tools produce as their output documents that conform to PMML. These can then be visualised, run, or combined with other models as appropriate, all within The Data Miner's Arcade environment.

To What Extent can Data Mining be Proceduralised

Graham J. Williams
Panel Discussion
Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD-98)
Melbourne, April 1998.

High Performance Data Managment Issues in Data Mining

Graham J. Williams
Presented to the Workshop on Parallel and Distributed Data Mining
Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD-98)
Melbourne, April 1998.

Mining the Knowledge Mine: The Hot Spots Methodology for Mining Large Real World Databases

[[pdf](#) [ps](#) [ps.gz](#)]

Graham J. Williams and Zhexue Huang
in *Advanced Topics in Artificial Intelligence*
Lecture Notes in Artificial Intelligence
Volume 1342, Pages 340--348
Springer-Verlag, 1997

Abstract

As databases grow in size and complexity the task of adding value to the wealth of data becomes difficult. Data mining has emerged as the technology to add value to enormous databases by finding new and important snippets (or nuggets) of knowledge. With large training sets, however, extremely large collections of nuggets are being extracted, leading to much "fools gold" amongst which to fossick for the real gold. Attention is now being directed towards the problem of how to better focus on the most precious nuggets. This paper presents the hot spots methodology, adopting a multi-strategy and interactive approach to help focus on the important nuggets. The methodology first performs data mining and then explores the resulting models to find the important nuggets contained therein. This approach is demonstrated in insurance and fraud applications.

PEPNet: Parallel Evolutionary Programming for Constructing Artificial Neural Networks

[[pdf](#) [ps](#) [ps.gz](#)].

Gerrit A. Riessen, Graham J. Williams, and Xin Yao
Sixth Annual Conference On Evolutionary Programming (EP97)
Indianapolis

Abstract

This paper presents a description of an evolutionary artificial neural network algorithm, EPNet and its extension taking advantage of a High Performance Computing Environment. PEPNet, Parallel EPNet, implements four forms of parallelism and this paper describes two of those parallelisms. Experimental studies have shown promising results with better time and prediction performance.

A Case Study in Knowledge Acquisition for Insurance Risk Assessment using a KDD Methodology

[[pdf](#) [ps](#) [ps.gz](#)]

Graham J. Williams and Zhexue Huang
Pacific Rim Knowledge Acquisition Workshop (PKAW96)
Sydney

Abstract

We describe some initial experiences in dealing with the task of acquiring knowledge where a very large collection of case histories is available. A Knowledge Discovery in Databases (KDD) approach is taken. KDD is the process of extracting novel information and knowledge from large databases, consisting of many interacting stages performing specific data manipulation and transformation operations with an information flow from one stage onto the next (and usually with feedback into previous stages). We characterise our experiences of this process for the task of acquiring knowledge for the domain of motor vehicle insurance premium setting for NRMA Insurance Limited.

Parallel Decision Tree Induction

Graham J. Williams
CSIRO DIT Data Mining Technical Report TR-DM-96024

Abstract

Knowledge discovery in databases (or KDD) and its associated data mining technologies are making enormous demands on traditional machine learning and statistical algorithms. KDD often deals with extremely large databases, often in sizes measured in terms of gigabytes rather than megabytes. Traditional machine learning and statistical techniques begin to be stretched beyond their capabilities when the data sizes reach many thousands of records. In this paper I review our work in dealing with very large datasets in the context of traditional decision tree induction algorithms (ID3 and C4.5). MIL (Williams 1988, Williams 1990), for Multiple Inductive Learning, is a system for inducing multiple decision trees in parallel, transforming those trees to rules, and then intelligently merging the resulting rule sets into a unified knowledge base. Our efforts to parallelise the decision tree induction algorithm for the Fujitsu AP1000 and the Thinking Machine Corporation's CM-5 high performance computers are also reviewed.

PEPNet: Parallel Evolutionary Artificial Neural Networks (Poster)

[[pdf](#) [ps](#) [ps.gz](#)]

Gerrit Riessen, Xin Yao, Zhexue Huang, Peter Milne, and Graham Williams
Fifth Australian Conference on Neural Networks (ACNN96)

Abstract

Artificial Neural Networks (ANNs) provide an important classification tool for Knowledge Discovery in Databases (KDD). Unfortunately ANNs require considerable time to train, particularly when large datasets are involved. Training time is also adversely affected when the characteristics of the dataset are not consistent with the structure of the ANN. In developing ANNs there are no hard and fast rules for determining the structure of the network. Evolutionary Artificial Neural Networks (EANNs) take advantage of evolutionary search techniques to address some of the problems associated with developing optimal ANNs. EANNs dynamically modify the structure of the ANNs on the basis of performance. EPNet (Yao and Liu 1996) is a serial algorithm which adopts these ideas to produce efficient ANNs. Such techniques produce greater accuracy in the networks, however at the expense of extra computational and storage requirements. Our work focuses on PEANNs, Parallel Evolutionary Artificial Neural Networks. PEANNs have the potential to produce accurate networks in significantly less time than serial EANNs using larger datasets. A parallel implementation of EPNet, called PEPNet, is being developed to explore this hypothesis.

KDD for Insurance Risk Assessment

Graham J. Williams and Zhexue Huang
March 1996
CSIRO DIT Data Mining Technical Report TR-DM-96014

Abstract

Insurance is a business of risks. Identifying and understanding areas of risk is an important task performed by an insurer. An assessment of risk is used to set the appropriate premium for insurance policies. This paper describes a KDD exercise which uses decision tree techniques to identify significant areas of risk within an insurance portfolio. The real world dataset used contains information about policies and insurance claims on those policies. Decision trees can be

constructed to identify and describe areas of high risk which are then evaluated, as a separate exercise, in terms of claim frequency and claim costs. The paper stresses the idea of interactive post-processing, or evaluation, of the patterns that are illuminated by traditional data mining tools.

Modelling the KDD Process

Graham J. Williams and Zhexue Huang

February 1996

CSIRO DIT Data Mining Technical Report TR-DM-96013

Abstract

Knowledge Discovery in Databases (KDD) is the process of extracting novel information and knowledge from large databases. This process consists of many interacting stages performing specific data manipulation and transformation operations with an information flow from one stage onto the next (and often back into previous stages). The process can be very complex and may exhibit much variety in the context of the variety tasks undertaken within KDD. In this paper we characterise our experiences of the KDD process and formalise its key elements in a model. A case study of insurance risk analysis for policy premium setting is used to illustrate the process and the model. The model provides a framework for comparing and differentiating various approaches to KDD.

Inducing and Combining Multiple Decision Trees

[280K gzip Postscript]

Graham J. Williams

PhD Thesis, Australian National University,
Canberra, Australia, 1990

Abstract

Most activities in our daily life require us to make decisions, many subconsciously. Knowledge is the key to correct decision making. Its representation and use by machine has been a major goal throughout the history of computing machinery. Learning is one of the most important components of intelligence and is a crucial aspect of knowledge-based systems. The research reported on here focuses on the acquisition of decision trees and their transformation to rules. A well-established practical tool for machine learning (ID3) is used as a basis for an approach to building, and then combining, multiple decision trees.

Combining Decision Trees: Initial results from the MIL algorithm

Graham J Williams

Artificial Intelligence Developments and Applications

edited by J. S. Gero and R. B. Stanton

Elsevier Science Publishers

1988, Pages 273-289

Some Experiments in Decision Tree Induction

Graham J Williams

Australian Computer Journal
1987, Volume 19, Number 2, Pages 84-91

Spatial Reasoning

Design of Decision Support Systems as Federated Information Systems

D. J. Abel, Kerry Taylor, Gavin Walker, and Graham Williams
Decision Support Systems for Sustainable Development
Edited by Kersten, Mikolajuk, and Yeh
Kluwer Academic Publishers, 1999

Templates for Spatial Reasoning in Responsive GIS

[8K gzip Postscript, first two pages only]

Graham J Williams
International Journal of Geographical Information Systems
1995, Volume 9, Number 2, Pages 117-131

Abstract

Responsive geographical information systems (GIS) address the needs of the decision maker working in a spatially oriented environment where data is regularly updated, where the data is often voluminous, incomplete, and noisy, and where timely decisions must be made. Such environments stretch the capabilities of traditional GIS. A responsive GIS must play a more active role in the support of the decision maker. This paper introduces the concept of a responsive GIS and demonstrates the integration of *artificial intelligence* techniques and *object-oriented database* technology to provide such active support. Expert knowledge, represented as Templates, can have both spatial and temporal components, and remains within the GIS framework rather than providing separate, and often disjoint, GIS and Expert System modules.

Representing Expectations in Spatial Information Systems

Graham J Williams and Steve G. Woods
Advances in Spatial Databases: Third International Symposium, SSD '93
Edited by D. J. Abel and B. C. Ooi
Lecture Notes in Computer Science, Volume 692, Springer-Verlag, 1993

GEM: A Micro-Computer Based Expert System for Geographic Domains

Graham J. Williams, J. Richard Davis and Paul M. Nanninga
Proceedings of the Sixth International Workshop and Conference on Expert Systems and Their Applications
Avignon, France, 1986.

The Design of Expert Systems for Environmental Management

J. Richard Davis, Paul M. Nanninga and G. J. Williams
Readings in Australian Geography
Proceedings of the 21st IAG Conference
Perth, Australia, 1988

Geographic Expert Systems for Resource Management

J. Richard Davis, Paul M. Nanninga and G. J. Williams
Proceedings of the First Australian Conference on Applications of Expert Systems
Sydney, Australia, 1985

Databases and Decision Support Systems

The Design of Decision Support Systems as Federated Information Systems

D. J. Abel, K. L. Taylor, G. C. Walker, and G. J. Williams
Proceedings of the Decision Support Systems for Developing Nations, Taiwan. December 1995

The Virtual Database: A Tool for Migration from Legacy LIS

D. J. Abel, B. C. Ooi, R. A. Power, K.-L. Tan, G. J. Williams, and X. Zhou
In Proceedings of the 22nd Annual Conference of the Australian Urban and Regional Information Systems Association AURISA '94
Sydney, Australia, 1994

The Virtual Database

David J. Abel, Beng Chin Ooi, Robert A. Power, Kian-Lee Tan, Graham J. Williams, and Xiafang Zhou
In Proceedings of the 21st Annual Conference of the Australian Urban and Regional Information Systems Association AURISA '93
Sydney, Australia, 1993

Knowledge Representation

FrameUp: A frames formalism for expert systems

Graham J Williams
Australian Computer Journal
1989, Volume 21, Number 1, Pages 33-40

Abstract

This paper presents an introduction to frames-based representation schemes for use in the construction of rule-based expert systems. The features that are relevant to such expert systems are discussed, followed by an example of the type of rule application mechanism that the system implements. Advantages of such a system are discussed.

This document is maintained by Graham Williams.

Last Modified 2002/04/16 14:24:22 Graham.Williams@cmis.csiro.au